data-science-guide

Release 0.1

Divye Gupta

Sep 03, 2023

CONTENTS

Data Science Guide is a Data Science Application using Python learning portal for beginer to intermediate data science enthusiasts who want to apply data science to real world busiess problems.

It pulls data from open source databases and presents the code in functional format to demonstrate how the concepts of machine learning can be applied to business problems and offers a simple guide to understand how data science can be leveraged by product managers, data analysts, business intelligence engineers, business analysts, consultants and data scientists using python.

Check out the *Why this Document?* section for further information, including how to understand and navigate the document to maximize learning.

Note: This project is under active development. Please contact the author for any questions or comments.

ONE

CONTENTS

1.1 Why this Document?

The data science guide is created by Divye Gupta, a Data Science enthusiast. Motivation behind this guide came after working on series of projects that belong to different data science disciplines and often it is impossible to remember all the syntax in python to do the analysis. This guide serves as a single source of truth for refering to python code and contains interpretations of results. Using this guide, anyone can quickly brush up their knowledge about key concepts of data science and apply them in their day to day solving of business problems.

Data Science is a very broad (science breath) and deep (science depth) field of science that deals with finding patterns in data and building mechanisms using programmatic interventions to maximize business outcomes. Although, the field has recently exploded (2014 onwards) to broader audience, research has been going on since mid 20th century. Disciplines of data science can be divided into the following broader categories:

1.1.1 Classical Learning

Classical Learning is the first innovation done in the field of data science. Also known as statistical learning, this field of science deals with linear models like Linear Regression, Logistic Regression, statistical feature engineering to make recommendations and communicate linear relationships between features and target variable

1.1.2 Machine Learning

Visit these documents to understand how you can connect science to business problems and implement the solutions in python with little to no programming background. Before, we start applying models and algorithms to a dataset, 60% of the time is devoted to preprocessing and visualizing the data. Check out the visualizer to see what functions you can use to quickly plot some graphs and communicate insights to the business to gain high level idea about the current situation in the business.